

# **Ciência da Computação – Universidade Federal do Agreste de Pernambuco (BCC-UFAPE) – Garanhuns – PE – Brasil**

## **Projeto da Disciplina**

### **Disciplina Optativa: Fundamentos de Ciência de Dados**

#### **Professores Responsáveis**

Ryan Azevedo & Assuero Ximenes

{ryan.azevedo, assuero.ximenes}@ufape.edu.br

## **Parte 1 – Equipes NBA**

Todo o projeto deve ser implementado/desenvolvido em Python. A parte de visualização de Dados deve ser em Python ou em Python e *Java Script*, os gráficos podem ser gráficos dinâmicos ou estáticos. Podem usar funções do Python para implementar as estatísticas e os algoritmos necessários.

### **Engenharia de Dados (Para Todo o Projeto)**

A primeira etapa do projeto consiste na Engenharia de Dados. As atividades que devem ser realizadas nessa etapa são:

- Extração dos dados dos jogos e jogadores (Apenas do seu time e dos seus Jogadores) da NBA temporada 23-24 (passada) e 24-25 (atual).
- Nas bases [isso pode ser e [deve ser] feito durante o ciclo de desenvolvimento] verificar dados ausentes, valores redundantes, ruidosos, inconsistentes, enviesados e valores outliers caso existam.
- Transformar os dados, verificar tipos, converter dados qualitativos para quantitativos, excluir ou não utilizar colunas do *Dataset* que não impactam [ainda] no projeto [Ex.: salário do jogador], normalizar valores numéricos quando necessário, preencher dados ausentes com dados reais e disponíveis pela entidade/empresa/corporação NBA [no caso de placar de jogos não extraídos automaticamente].
- Como *amostras de dados* vamos utilizar todos os dados do *Dataset* (100% dos dados que nos interessa). [Não estamos preocupados com desempenho computacional].
- O procedimento de separação [Treinamento e teste] do conjunto de dados será realizado apenas durante o desenvolvimento dos modelos preditivos e estatísticos que usaremos. Aqui usaremos técnicas automáticas para evitar o *data leakage*.

Com os dados coletados, passamos para a fase de desenvolvimento das funcionalidades relacionadas ao time/equipe.

## Funcionalidades/Requisitos – Equipe/Time<sup>1</sup>

O sistema deve proporcionar aos *stakeholders* um conjunto de funcionalidades que os auxiliem no processo de tomada de decisão. Os dados utilizados serão os da **atual temporada (24-25)** da NBA e da temporada passada (**23-24**). O sistema deverá atender aos Requisitos Funcionais (RF) definidos e detalhados neste documento de especificação.

### Requisitos Funcionais:

**RF1** – Listar todos os times da NBA [Dependendo de “onde esteja extraíndo” os dados, existe um Id, exiba também o Id de cada time] agrupados por Conferência (Vide Tabela 1). Obs.: Não considerar a divisão pertencente ao time.

Conferência Leste	Conferência Oeste
-------------------	-------------------

Tabela 1. Lista de Times por Conferência

**RF2** – Apresentar a classificação <sup>2</sup>atual dos times [Agrupados por Conferência]

### Para sua Equipe/Time, faça:

**RF3** – Apresentar o total de vitórias e derrotas do time, separados por partidas jogadas em casa (mandante) e fora de casa (visitante), conforme Tabela 2;

Total de Vitórias	Total de Vitórias em Casa	Total de Vitórias fora de casa	Total de Derrotas	Total de Derrotas em Casa	Total de Derrotas Fora de casa
-------------------	---------------------------	--------------------------------	-------------------	---------------------------	--------------------------------

Tabela 2. Resultados do Time

**RF4** – Apresentar o total dos dados do seu time [temporada 23-24 e temporada atual], conforme Tabelas 3 e 4;

Total de Pontos por Jogo	Total de Assistências por Jogo	Total de Rebotes por Jogo	Total de Cestas de 3 Pontos Convertidas	Derrotas em Casa	Derrotas Fora de casa
--------------------------	--------------------------------	---------------------------	---	------------------	-----------------------

Tabela 3. Total de Dados

**RF5** – Apresentar a divisão de alguns dados da Tabela 3, conforme Tabelas 4

Total de Rebotes	Total de Rebotes Ofensivos	Total de Rebotes Defensivos	Total de Pontos	Total de Cestas de 2 Pontos	Total de Cestas de 3 Pontos	Total de Lance Livre
------------------	----------------------------	-----------------------------	-----------------	-----------------------------	-----------------------------	----------------------

Tabela 4. Divisão do Total de Dados

**RF6** – Apresentar os dados referentes a performance defensiva do time [temporada 23-24 e temporada atual], conforme Tabela 5.

<sup>1</sup> Todos os alunos já sabem seus times e podem verificar nos Apêndice 1 e 2 deste documento.

<sup>2</sup> Posição atual no campeonato

Total de Roubos de bola	Total de Rebotes Defensivos	Total de Tocos por jogo	Total de Erros por jogo	Total de Faltas por jogo
-------------------------	-----------------------------	-------------------------	-------------------------	--------------------------

Tabela 5. Total de Dados Defensivos

**RF7** – Apresentar os jogos do seu time [temporada 23-24 e temporada atual], conforme Tabela 6.

Data do Jogo	Adversário	Vitoria ou Derrota	Casa ou Fora	Placar
--------------	------------	--------------------	--------------	--------

Tabela 6. Total de Dados Defensivos

**RF8** – Apresentar gráficos de desempenho do seu time [temporada 23-24 e temporada atual] para compor o *Dashboard* do projeto:

- Gráfico de Barras Empilhado para Vitórias [Cor verde] e Derrotas [Cor vermelha].
- Gráfico de Barras Agrupado para Vitórias em casa [Cor verde], Vitórias fora de casa [Cor azul], Derrotas em casa [Cor vermelha] e Derrotas fora de casa [Cor marrom].
- Gráfico Histograma para exibir a frequência de vitórias e derrotas do time.
- Gráfico de Setor [Pizza] para o percentual de para Vitórias em casa, Vitórias fora de casa, Derrotas em casa e Derrotas fora de casa.
- Gráfico de Radar exibindo a média de pontos marcados e sofridos nos jogos em casa e fora de casa.
- Gráfico de Linhas exibindo a sequência de vitórias e derrotas ao longo da temporada.
- Gráfico de Dispersão exibindo equipes e a média de pontos marcados e sofridos durante a temporada
- Apresente um gráfico da sua escolha para exibir os dados referentes a Tabela 5 do **RF6** e a Tabela 6 do **RF7**.

**RF9** – Para cada tabela e/ou requisito que exiba dados gerados pelo sistema será necessário que os dados sejam salvos em um arquivo *csv*.

**RF10** – Para cada gráfico gerado pelo sistema será necessário que sejam exibidos em formato HTML e abertos no *Browser*.